学校编码：10384　　　　分类号　　密级
学号：X200423006　　　　UDC

厦门大学

# 硕 士 学 位 论 文

# 极值分布模型在极端气温变化领域上的初步研究

## Preliminary study of extreme value distribution model in the areas of extreme temperature changes

徐 春 香

指导教师姓名：黄　荣　坦

专 业 名 称：基 础 数 学

论文提交日期：2009 年 7 月

论文答辩时间：2009 年 9 月

学位授予日期：2009 年　月

答辩委员会主席：

评　　阅　　人：

2009 年　月

# 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外,该学位论文为(　　　　　　　　　　　　　　　　)课题(组)的研究成果，获得（　　　　　　　　　　）课题（组）经费或实验室的资助，在（　　　　　　　　　）实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。)

声明人（签名）：徐春香

**2009** 年 9 月 14 日

# 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　　年　　月　　日解密，解密后适用上述授权。

（　　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：徐春香

2009 年 9 月 14 日

# 摘 要

极值分布模型是一个研究极值现象和极值随机变量的非常有效的工具。极值理论主要以极值为研究对象，它注重分布的尾部。因此，越来越多的人认识到极值理论在极端事件运用中的巨大潜力，特别指出的是极值理论是一种模拟数据分布尾部的理论，所以可以应用于极值数据的预测。

本文首先分析了在我们周围极值现象广泛存在，对自然界的平衡以及人类社会的发展都具有重要的影响。充分研究极值现象，有效掌握其规律并积极采取相应的措施，可以帮助人们在遇到突发事件时降低风险减少损失。然后系统介绍极值理论知识，重点介绍了一维极值分布模型及广义帕雷托分布模型。最后利用永泰县 1997～2008 年气象观测基础数据,应用极值模型建立永泰县夏季、冬季极端最高气温的概率分布模型, 发现模型能够较好地反映这些数据的特征。依据极值分布模型可计算永泰县某年内夏、冬季可能出现的极端最高气温,这为农作物品种的引进及农作物栽培时防高温抗干旱提供了叁考依据。

**关键词**：极端气温；极值模型；帕雷托分布

# Abstract

Extreme value distribution model is a very effective tool for studying the phenomenon of extreme value and extreme value of the random variable. Extreme Value Theory mainly studys extreme value, which focus on the tail of the distribution. As a result, more and more people recognize the great potent that extreme value theory used in extreme events, in particular, the extreme value theory is a simulation of the tail of the distribution of information theory, the data can be used in extreme forecasts.

This paper first analyzes the widespread phenomenon of the extreme around us. The extreme has significant impacts on balance of nature as well as the development of human society. Fully study the extreme situation,effectively master the law and actively take corresponding measures can help people reduce the risk of losses in the face of unexpected events. Then fully introduces Extreme Value Theory, focusing on the introduction of one-dimensional model of extreme value distribution and the generalized Pareto distribution model. Finally on the basis of the basic meteorological data from 1997 to 2008 of YongTai County, the paper uses extreme value theory model to set up the probability distribution model of extreme maximum temperature in summer and winter of YongTai. It finds that the model can better reflect the characteristics of these data. According to the extreme value distribution model,we can caculate the extreme maxium temperature that may appears in summer and winter in YongTai some years.It provides a scientific basis for the introduction of new varieties and the growth of crops.

**Key words**：Extreme weather；Extreme value model；Pareto distribution

# 目　录

# Catalogue

# 第一章 引言

## 1.1 研究意义

天气和气候的极值事件往往给人类社会和生态系统带来严重的破坏性后果.例如:洪水，干旱，破坏性大风，热浪[1](Heat Wave)和严寒等。尤其对于农业，一次极值事件有可能造成农作物大规模减产。以极端气温为例，全球变暖所导致的极端温度变化，已在许多地区引起死亡率上升。尽管严寒出现次数减少，但是酷夏造成的死亡率上升幅度往往超过暖冬带来的死亡率下降幅度。因此，热浪在全球自然灾害中的排名为第七位，可见其影响之大。此外，极端温度及其空间分布变化对农作物生长发育及最终产量的巨大影响，也是众所周知的，它们往往会导致农作物品种种植结构的改变，其危害性已使许多农业专家感到棘手。当然，极端事件有时也对人类和生态带来好处。比如:在沙漠的绿洲中生活的动植物其生存的水源主要依靠突发性的极端降水； 1997-1998 年，厄尔尼诺事件造成美国冬季少有的高温，用于取暖的能源消耗节省了近 56 亿美元,800 多人幸免于严寒带来的死亡；澳大利亚西北部沿海大部分降雨来自偶发性的热带气旋，没有这些降雨，那里的人类和生态会受到很大的破坏。因此，掌握极端天气和气候事件的产生、发展的演变规律，可以趋利避害，有效地利用自然，引起了实际应用者和研究者们的广泛关注，极值理论的研究得到了进一步的关注。

极植，本意是指稀有的、极端的、偏激的，或在人们经验范围内很少出现或发生的事件。从统计学意义上讲，是指某一时期的随机过程的最大值和最小值，通常位于分布的尾部。从实践的观点看，研究分布的尾部的重要意义在于能帮助我们估计极值的运动规律，一旦知道了尾部特征，我们就可以应用一定的工具进一步分析可能的极值运动。其最重要的意义在于能评估极端事件的风险，比如百年不遇的洪水、大地震、金融危机等。极值理论应用研究最早是在 20 世纪 30 年代,主要应用到材料科学、洪水分析、地震分析和降雨量分析等方面。其中,Gumbel 对极值理论的应用研究做出了重要贡献,他第一个将极值理论系统地应用到实践中，并引起了工程师和统计学家的注意[2,3]

永泰县位于福建中部,是戴云山脉自西向东延伸部分的丘陵地带。全县地势

高低不平，最高海拔1653m，最低海拔仅10m。境内群山林立，大樟溪自西向东，横贯中部，形成长廊式的谷地，是“八山一水一分田”的山区农业县，属亚热带季风气候区。极端最高气温是夏季农作物生长与分布的主要限制因素，如2003年的持续高温就造成了全县的农作物普遍受到不同程度的影响。而且，近几年来为了改变该县单一的农作物结构，提高经济效益和生态效益，开始引进多种农作物，但由于极端最高气温会高达41.1 ℃，使新引进的农作物遭受到严重的酷热灾难，严重影响其生长，造成了极大的损失。为此，如何有效预防农作物发生酷热灾难具有重要的现实意义。要做到有效地预防农作物发生酷热灾难，就必须认真做好抗旱耐高温工作、了解品种的耐温性、了解种植区内的品种在一个生长期内可能遇到的各种概率值的极端最高气温，提供不同生长期的极端最高气温值。所以，必须寻找一种能够描述该县极端最高气温的概率分布的数学模型，而极值理论为研究极端气温提供了一种很有实用性的概率分布模型。

## 1.2 研究现状

极值理论(Extreme Value Theory，以下简称 EVT)是次序统计学的一个重要分支，主要研究的是极值分布及其特性，尤其是分布的尾部特征。其研究方法和范围经历了很大的变化[4]。早期阶段，极值只是研究一系列独立同分布的随机变量中的极大值和极小值应该服从何种分布。Fisher-tippett LHC（1928）发表了第一篇关于极值极限的理论文章，给出 Fisher-tippett 定理，说明样本的极大值或极小值在样本大小趋向无穷时，其分布收敛到三种分布中的一种，大大简化了对极值极限性质的研究，把多种分布函数按尾部性质分成相当简单的三类，从而建立了这门理论的基础。Gnedenko BV（1943）第一次给出了 Fisher-tippett 定理的严格数学证明。1950 年以英国的统计学者为主展开了以一种形式表现三种类型极值分布的广义极值分布的研究，使叁数估计问题变得简单，因不再需要从三种模型中选出一种建模而避免了程序的繁琐。Gumbei（1960）将极值理论应用到具体的统计问题上，并提出了现在称之为区组法的统计方法，即将按照事先定好的长度分成很多区间，然后从每个区间中选出一个最大值或最小值来进行建模。但是，从大量数据中仅选用极值会损失其他大量数据所含的信息。Pickands IIIJ

（1971）提出了新的极值研究方法，即选取某个界限以上的数据进行分析的方法称之为 POT 法。其分布函数的极限形式趋向广义帕雷托（Pareto）分布（GPD）。他的工作加深了对极值概念及其内在涵义的理解，即不仅仅最大值最小值是极值，而且在某个很大门限值以上的数据都是极值，都有研究的必要，使预测极值事件时可以使用更多的数据。该工作把极值统计研究的焦点从极大值方法转移到门限值方法上来。

## 1.3 本文的结构安排

本文在国内外已有的研究基础之上，对永泰县日最高气温数据进行较为详细的分析预测研究，对几种模型的应用及结果进行讨论。

第一章：简单介绍了极值理论的发展状况以及研究极值随机现象的重要性，并从永泰县较高的日气温影响了农作物生长、造成农作物品种相对单一的现状来说明论文研究的必要性。

第二章：系统地阐述了极值理论：一维极值分布模型、广义帕雷托分布(GPD)模型。

第三章：极值模型在夏季极端气温变化领域上应用。

第四章：极值模型在冬季极端气温变化领域上应用。

第五章：结论及研究趋势。

# 第二章 极值模型

极值，从统计学角度看，是指某一时期的随机过程的最大值和最小值，通常位于数据分布的尾部。极值分布是指观测值中极大值或极小值的概率分布。极值理论的对象是极端值，主要研究极端值的分布及其特征，尤其是分布的尾部特征，根据分布的尾部特征可以进一步预测可能的极值运动。也就是说极值理论是一门用来预测异常现象或者小概率事件风险的模型技术，它具有超越样本数据的估计能力，并可以准确地描述分布尾部的分位数。

## 2.1 一维极值分布模型

### 2.1.1 一维极值分布模型的初步理论

一维极值分布模型是一种与经典的统计模型的视角和侧重点都有很大不同的模型，它不考虑样本母体的中央分布，而将注意力放在两侧（tail）。

用 $X_1, X_2, \cdots X_n$ 表示独立同分布的随机变量，$X_i$（i=1，2，…n）的分布函数 F（x）未知。$X_i$ 可表示一个资产的收益率、信用风险值、市场风险值、大灾难保险索赔值或与特大自然灾难相联系的一些数据。这篇文章中 $X_i$ 表示一列正的数据。定义 $X_{(1)}=\min\limits_{1\le j\le n}\{X_j\}$，$X_{(n)}=\max\limits_{1\le j\le n}\{X_j\}$，分别称为样本极小值和样本极大值，统称为样本极值(为方便起见，样本极大值通常被记为 $M_n$)，它们的分布称为极值分布。由极值构成的样本数据具有独立同分布的特点。

设 $F_1(x)$ 为极小值分布函数，$F_n(x)$ 为极大值分布函数，则总体分布与它们之间有如下关系:

$$
\begin{aligned}
F_1(x) &= \Pr[X_{(1)}\le x]=1-\Pr[X_{(1)}>x] \\
&= 1-\Pr(X_1>x,\ X_2>x,\ \ldots,\ X_n>x) \\
&= 1-\Pr(X_1>x)\cdot\Pr(X_2>x)\cdot\ldots\cdot\Pr(X_n>x) \\
&= 1-[1-F(x)]^n
\end{aligned}
$$

$$
\begin{aligned}
F_n(x) &= \Pr[X_{(n)}\le x]=\Pr(X_1\le x,\ X_2\le x,\ \ldots,\ X_n\le x) \\
&= \Pr(X_1\le x)\cdot\Pr(X_2\le x)\cdot\ldots\cdot\Pr(X_n\le x) \\
&= [F(x)]^n
\end{aligned}
$$

就是说由以上两式，可以从总体分布得到极值分布。但大多数情况下，总体分布 F（x）是未知的，这就导致了依据渐近理论的方法，人们常用的是 $n\to\infty$ 时所得到的极值渐近分布。我们主要讨论 $X_{(n)}$ 的性质。

显然，当 $n\to\infty$ 时，$\Pr[X_{(n)}\le x]=[F(x)]^n$ 是一个退化的分布函数：

$[F(x)]^n=\begin{cases}0, & 0\le F(x)<1\\ 1, & F(x)=1\end{cases}$，没有实际研究意义。为此，通过对 $X_{(n)}$ 进行线性变换 $X^*_{(n)}=\dfrac{X_{(n)}-\beta_n}{\alpha_n}$ 解决，其中 $\{\alpha_n>0\}$ 和 $\{\beta_n\}$ 为适当的常数序列。$X^*_{(n)}$ 的分布函数成为一元极值分布模型的研究对象。

### 2.1.2 一维极值分布模型的分类

极值随机变量 $X^*_{(n)}$ 的所有可能形式的分布函数，是由费舍与逊皮特（Fisher-Tippett）通过一个定理给出[5]。Fisher-Tippett 定理(1928)是极值理论的核心，这个理论主要说明了极值分布的收敛特性。

定理 2.1[Fisher-tippett theorem]:

存在两个实数序列 $\{\alpha_n\}$ 和 $\{\beta_n\}$（其中 $\alpha_n>0$），满足

$$\lim_{n\to\infty}P\left(\frac{X_{(n)}-\beta_n}{\alpha_n}\le x\right)=H(X)\quad(x\in R)$$

其中 H（x）为非退化的分布函数，则 H（x）是下列三种类型的极值分布函数之一：

·类型 I：ξ>0,Frechet,其分布函数为：$\Phi_\xi(x)=\begin{cases}0, & x\le 0\\ \exp(-x^{-\xi}), & x>0\end{cases}$

·类型 II：ξ=0,Gumble,其分布函数为：

$$\Lambda(x)=\exp[-\exp(-x)],\quad x\in R$$

·类型 III：ξ<0,Weibull ,其分布函数为：$\Psi_\xi(x)=\begin{cases}\exp[-(-x)^{\xi}], & x<0\\ 1, & x\ge 0\end{cases}$

其中 $\xi$ 是形状参数。

这个定理描述了无论原始的分布函数 F(x)是什么形式，线性变换后的样本极大值，都依概率收敛于一个具有以上三种分布函数之一的随机变量。因此，定理

2.1 在极值理论体系中处于核心地位，为进一步的研究提供了坚实的基础。

### 2.1.3 一维广义极值分布模型

一维极值分布模型的一个实践难点是，对于获取的极值数据无法直接判断究竟属于三种分布类型的哪一类。为了便于进行统计推断，Uon Mrses(1954)与Jenkinson(1955)提出的广义极值分布(GEV)是极值的近似分布。

广义极值分布函数为：

$$H_{\xi,\mu,\sigma}(x)=\begin{cases}\exp[-(1+\xi\frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}], & \xi\neq 0, 1+\xi\frac{x-\mu}{\sigma}>0\\ \exp[-\exp(-\frac{x-\mu}{\sigma})], & \xi=0, -\infty<x<+\infty\end{cases}$$

其中，$-\infty<\mu<\infty$ 为位置参数，$\sigma>0$ 为尺度参数，$-\infty<\xi<\infty$ 为形状参数。数据 $X_i$ 的分布函数 F（x）的尾行为决定了广义极值分布 H（x）的形状叁数 ξ。如果 F（x）的尾部是指数衰减的，则 H（x）是 Gumble 类型且 ξ=0。Gumble 族包含了如正态分布、对数正态分布、指数分布和 Gamma 分布等薄尾分布；如果 F（x）的尾部是幂函数衰减的，则 H（x）是 Frechet 类型且 ξ>0，Frechet 族包含了如 Pareto 分布、Cauchy 分布、学生-t 分布等重尾分布；如果 F（x）的尾部是有限的，则 H（x）是 Weibull 类型且 ξ<0，Weibull 族包含了如均匀分布、β-分布等。

对应的广义极值分布的密度函数为：

$$h(x)=\begin{cases}\exp\{-[1+(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}}\}\frac{1}{\sigma}[1+\xi(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}-1}, & \xi\neq 0, 1+\xi\frac{x-\mu}{\sigma}>0\\ \exp\{-\exp[-(\frac{x-\mu}{\sigma})]\}\frac{1}{\sigma}\exp[-(\frac{x-\mu}{\sigma})], & \xi=0, -\infty<x<\infty\end{cases}$$

图 1 是广义极值密度函数图 $h_{0,1,\xi}(x)$，它分别考虑了形状参数 $\xi$ 取正值、零和负值来说明广义极值分布的特征，令 $\xi=0.4$ 的 Frechet 分布和 ξ=0 的 Gumbel 分布以及 $\xi=-0.4$ 的 Weibull 分布。我们可以清楚地看出 Frechet 分布左尾是一个幂函数衰减、左边界是厚尾分布； Gumbel 分布左尾是指数衰减的、左

边界是薄尾分布；weibull 分布的下端点是有限的。

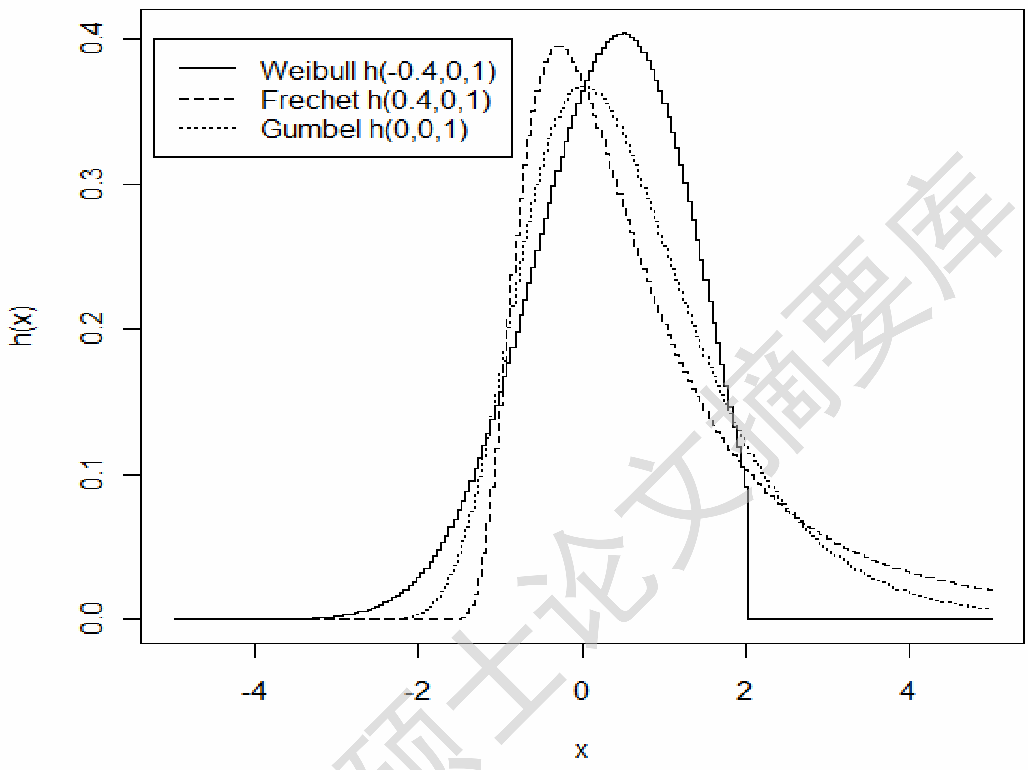


图 1 是广义极值密度函数图 $h_{0,1,\xi}(x)$

### 2.1.4 参数估计

参数估计是极值分布模型的核心问题之一，常用的方法有极大似然法、矩法、概率权值矩估计和回归法等。极大似然估计充分利用了总体分布函数表达式所提供的信息，其统计思想符合人们的认识和经验，因而有很多优良的性质，如极大似然估计的渐进分布是正态分布等。正因为如此，极大似然估计法是最重要和最好的方法之一，是最常用的估计方法。但是这个过程中需要数值求解非线性方程组，运算量较大，应用起来不方便。Jenlinson（1969），Hosking（1985），Macleod（1989）等给出了具体的算法和步骤。

**(1)极大似然估计**

极大似然估计是统计学中最重要的、也是应用最广泛的方法之一。其思想最初出现在 Fisher1912 年一项工作中，之后他又对相关的性质作了理论分析，是 Fisher 对统计学的重大贡献之一。